J. Biosoc. Sci., (2018) **50**, 868–869,
doi:10.1017/S0021932018000172 First published online 17 Jul 2018

Debate

IGNORING PSYCHOMETRIC PROBLEMS IN THE STUDY OF GROUP DIFFERENCES IN COGNITIVE TEST PERFORMANCE

JELTE M. WICHERTS[1]

Department of Methodology and Statistics, Tilburg University, Tilburg, the Netherlands

Summary. In a recent study, te Nijenhuis *et al.* (2017) used a version of Jensen's method of correlated vectors to study the nature of ethnic group differences on Raven's Progressive Matrices test. In this comment, the author points out that this method has been shown to be psychometrically inappropriate in studying group differences in performance on dichotomous (correctly or incorrectly scored) items. Specifically, the method uses item statistics like the item-total correlation that necessarily differ across groups differing in ability and employs a linear model to test inherent non-linear relations. Wicherts (2017) showed that this method can provide correlations far exceeding $r = 0.44$ in cases where the group differences cannot possibly be on *g* because the items measure different traits across the groups. The psychometric problems with their method cast serious doubts on te Nijenhuis *et al.*'s conclusions concerning the role of *g* in the studied group difference in cognitive test performance.

In their recent article on ethnic group differences on Raven's well-known Progressive Matrices test, te Nijenhuis *et al.* (2017) refer to criticisms that I levelled against work on ethnic group differences in cognitive ability (IQ) tests. The criticism that they (indirectly) referred to concerned studies (Templer & Arikawa, 2006; Kanazawa, 2008) that denied several well established facts. Specifically, several methods in these studies denied the Flynn Effect, climate change since the last ice age, global migration since the 1500s and the earth not being flat (Wicherts *et al.*, 2010), to mention some of these studies' more egregious errors. But my critique of those earlier studies on national IQ is hardly relevant for the psychometric analyses used by te Nijenhuis *et al.* to study why groups show varying performance on the Raven's test. Unfortunately, their analyses too, are rather lethally flawed.

[1] Email: j.m.wicherts@uvt.nl

Wicherts and Johnson (2009) already highlighted psychometric problems with Jensen's (1998) method of correlated vectors applied to dichotomous (correctly or incorrectly scored) item scores. Notably, the item statistics used in this method, like the item-total correlation, have long been known (e.g. Ferguson, 1941) to be dependent on the ability of the group. Another problem is that the relations between item statistics are inherently and complexly non-linear because of the statistics being bounded by 1. After noting that te Nijenhuis and colleagues had ignored Wicherts and Johnson (2009) in earlier works using the method of correlated vectors, I wrote a more elaborate critique (Wicherts, 2017). I offered the latter manuscript to te Nijenhuis *et al*. when I reviewed an earlier version of their current paper for a psychometric journal, but they declined to read it as it was still under review at the time. Their manuscript is now published in the *Journal of Biosocial Science* but continues to ignore the psychometric problems with the method of correlated vectors.

Wicherts (2017) extensively argued that the method of correlated vectors is patently incapable of highlighting whether or not *g* is indeed the source of group differences on the Raven's test. Even in the unlikely scenario that Raven's items measure only *g* and do so perfectly and in the same manner in both groups (a hypothetical case in which Spearman's hypothesis is true), the method can yield basically any correlation, even negative ones. Moreover, Wicherts (2017) showed empirically that the method is incapable of detecting instances in which Spearman's hypothesis is not true because studied items measure different traits across groups. Crucially, in these instances, the method can yield correlations between the vectors that far exceed the correlation of 0.44 found by te Nijenhuis *et al.* (2017). This result casts strong doubt on their conclusion that such a correlation corroborates that the differences can be (more or less) attributed to *g*. The method of correlated vectors applied to item scores lacks diagnostic value in studying whether or not group difference are caused by *g* (Wicherts, 2017) and so I suggest the authors apply established psychometric methods to their interesting data.

References

Ferguson, G. A. (1941) The factorial interpretation of test difficulty. *Psychometrika* **6**, 323–329.

Jensen, A. R. (1998) *The g Factor: The Science of Mental Ability*. Praeger Publishers, Westport, CT.

Kanazawa, S. (2008) Temperature and evolutionary novelty as forces behind the evolution of general intelligence. *Intelligence* **36**, 99–108.

te Nijenhuis, J., Batterjee, A. A., Van Den Hoek, M., Allik, J. & Sukhanovskiy, V. (2017) Spearman's hypothesis testing comparing Saudi Arabian children and adolescents with various other groups of children and adolescents on the items of the Standard Progressive Matrices. *Journal of Biosocial Science* **49**(5), 634–647.

Templer, D. I. & Arikawa, H. (2006) Temperature, skin color, per capita income, and IQ: an international perspective. *Intelligence* **34**, 121–139.

Wicherts, J. M. (2017) Psychometric problems with the method of correlated vectors applied to item scores (including some nonsensical results). *Intelligence* **60**, 26–38.

Wicherts, J. M., Borsboom, D. & Dolan, C. V. (2010) Why national IQs do not support evolutionary theories of intelligence. *Personality and Individual Differences* **48**, 91–96.

Wicherts, J. M. & Johnson, W. (2009) Group differences in the heritability of items and test scores. *Proceedings of the Royal Society: Series B* **276**, 2675–2683.